Characterizing the Distribution of Macronutrient Intake among U.S. Adults: A Quantile Regression Approach

Jayachandran N. Variyam, James Blaylock, and David Smallwood

Since the risk of dietary inadequacy or excess is greater at the tails of the nutrient intake distributions than at the mean, marginal effects of explanatory variables estimated at the conditional mean using ordinary least squares may be of limited value in characterizing these distributions. Quantile regression is effective in this situation since it can estimate conditional functions at any part of the distribution. Quantile regression results suggest that age, education, and income have a larger influence at intake levels where the risk of excess is greater compared with intake levels where the risk of excess is lower.

Key words: diet quality, health risk, heteroskedasticity, nutrition.

The growing evidence on the health effects of foods, nutrients, and other dietary components has heightened interest in the composition of U.S. food demand and supply (Kantor) and the quality of American diets and their determinants (Adelaja, Nayga, and Lauderbach; Bowman et al.; Chern; Gould 1996; Nayga). An important but often overlooked consideration when studying dietary intakes is that the risk of inadequacy or excess, and the risk of adverse health effects, is greater at the tails of the intake distributions than at the mean.

To illustrate this point, table 1 reports the mean and selected percentiles of the daily intake of four major macronutrients among men and women 18 years or older, obtained from the 1994–96 Continuing Survey of Food Intakes by Individuals (CSFII). Table 2 reports the recommended daily intakes for these same macronutrients. Clearly, most mean and median intakes are within the daily intake levels recommended by health authorities. For example, the recommended daily fat intake for men between age 19 and 50 is less than or equal to 96.7 grams of fat, based on a 2900-calorie diet. The mean and median intakes of total fat for men during 1994–96 are about 92 and 84 grams, well within this recommended level. However, intake at the 90th percentile is 146 grams, considerably above the healthful level. Similarly, for women, the cholesterol intake even at the 75th percentile, 273 milligrams (mg), is below the recommended daily intake of 300 mg, whereas at the 90th percentile the intake (389 mg) is well above the healthful level. Clearly, from a public health and nutrition policy perspective, characterizing the population at the tails of these dietary intake distributions (75th and 90th percentiles of saturated fat, for example) is of greater interest than studying those around the mean. Suppose we want to examine the difference in the intake of a nutrient between two demographic groups. If we limit our comparison to mean intakes, we would know only the *average* between-group difference over the whole range of the intake distribution. This implicitly assumes that the between-group difference in intake is the same along the whole distribution and neglects any pronounced difference at one tail or the other. From a policy perspective, the *location* of the intergroup difference along the whole distribution is arguably more important than the *average* between-group difference. For example, for a nutrient whose excessive intake is of concern, policy makers would be more concerned if the between-group difference is located predominantly at the upper end of the distribution as opposed to the lower end of the intake distribution.

The authors are agricultural economists with the Economic Research Service, USDA.

The authors thank Betsy Frazao and anonymous reviewers for helpful comments. The views expressed in this article are those of the authors and do not necessarily reflect the views of the Economic Research Service or the USDA.

Table 1. Distribution of Macronutrient Intake among U.S. Adults, 1994–96

Nutrient	Units	Percentile 10	25	50	75	90	Mean
Men							
Total fat	Grams	43.2	60.2	83.8	114.5	146.0	91.8
Saturated fat	Grams	13.3	19.4	27.9	38.8	50.9	31.0
Cholesterol	Milligrams	121.4	181.6	283.2	423.2	594.8	329.7
Fiber	Grams	8.0	11.6	16.2	22.9	29.8	18.1
Women							
Total fat	Grams	27.2	39.7	55.7	73.9	94.6	59.2
Saturated Fat	Grams	8.4	12.3	18.1	25.0	32.6	19.7
Cholesterol	Milligrams	71.6	113.2	176.5	272.9	389.3	209.6
Fiber	Grams	5.9	8.6	12.4	16.9	22.3	13.5

Note: Weighted estimates. Men: $N = 4882$; women: $N = 4714$.

Just as the mean may give an incomplete picture of the intake distribution, so might the conditional mean in linear regression models estimated by ordinary least squares (OLS). Linear regression models have been widely employed to determine marginal differences in nutrient intakes between population subgroups of interest to policymakers (Adelaja, Nayga, and Lauderbach; Chavas and Keplinger; Nayga). The OLS approach is satisfactory if the marginal effects of population characteristics are identical over the entire range of the dependent variable's conditional distribution. While this may be a reasonable assumption in many cases, it is unlikely to hold for dietary intakes because of the increasing risk of inadequacy or excess at the tails of nutrient intake distributions. Key variables such as education are likely to exert a different effect at the tails of the distribution compared with the mean, especially in the face of rapid increase in the flow of health information.

The early 1990s saw significant changes in nutrition promotion efforts (Davis and Saltos). The Food Guide Pyramid was introduced in 1992 and the Nutrition Labeling and Education Act went into effect in 1994. Economic models of information acquisition and use suggest that individuals with greater human capital may acquire, process, and use this type of new information with greater efficiency in their dietary decision-making (Ippolito and Mathios). Empirically, it is more enlightening to look for the evidence of such behavior at the tails of the intake distribution than at the mean. By focusing only on the conditional mean, the

Table 2. Recommended Daily Intakes of Selected Macronutrients

Sex and Age	Nutrient: Energy (Calories)	Total Fat[a] (Grams)	Saturated Fat[b] (Grams)	Cholesterol (Milligrams)	Fiber[c] (Grams)
Men					
19–20	2900	≤96.7	<32.2	≤300	Age + 5
21–24	2900	≤96.7	<32.2	≤300	33.4
25–50	2900	≤96.7	<32.2	≤300	33.4
51+	2300	≤76.7	<25.6	≤300	26.5
Women					
19–20	2000	≤66.7	<22.2	≤300	Age + 5
21–24	2000	≤66.7	<22.2	≤300	23
25–50	2000	≤66.7	<22.2	≤300	23
51+	1900	≤63.3	<21.1	≤300	21.9

Source: Lin, Guthrie, and Frazao, table 4, p. 6.
[a] Based on the recommendation that no more than 30% of total calories come from fat.
[b] Based on the recommendation that less than 10% of total calories come from saturated fat.
[c] For men and women 21 and above, the figures are based on the recommendation for a fiber intake of 11.5 grams per 1000 calories of energy.

OLS approach may, therefore, provide an incomplete picture of the factors promoting healthier dietary behaviors.

In this study, we employ quantile regression to better characterize macronutrient intake among U.S. adults by modeling parts of the conditional distribution of intakes other than the conditional mean. The quantile regression approach relaxes the assumption that the effects of explanatory variables are constant along the whole distribution of the dependent variable and allows such effects to vary over the entire range of dietary intake. This enables us to estimate the intake differentials of population subgroups at specific points of interest in the conditional distribution, such as the 90th percentile for saturated fat and the 10th percentile for fiber.

Empirical Approach

We estimated a set of regression equations that expressed the quantity of nutrients consumed by an individual as a function of his or her sociodemographic and anthropometric characteristics. These equations can be viewed as linear approximations of reduced form nutrient demand functions derived from Becker's household production model (Grossman; Strauss and Thomas). Alternatively, they can be viewed as reduced form demand functions derived from the characteristics model developed by Lancaster. In Becker's model, households gain utility from nonmarket commodities such as family members' health, which are in turn produced by the households by combining time, human capital, and purchased market goods. The solution to the household maximization problem subject to technology, income, and time constraints gives the demand functions for the final commodities (such as health), intermediate commodities (such as nutrients), and market goods (such as foods). In the Lancaster model, consumers maximize utility derived not from the goods themselves, but from the attributes of the goods they consume. Diets, comprising a combination of foods, embody positive attributes such as taste and essential nutrients, and negative attributes such as health risk associated with saturated fat and cholesterol (Chern).

In either case, sociodemographic factors enter into the demand functions because they influence a consumer's efficiency in producing and consuming health inputs or attributes. For example, educational attainment affects health production by raising technical and allocative efficiencies of input use (Grossman and Kaestner). Technical efficiency enables the more educated to produce a larger health output from a given level of health inputs. Allocative efficiency enables the more educated to acquire and use information about the true effects of inputs on health. Similarly, if the demand for health is inelastic and if health stock depreciates at an increasing rate with age, then health investment should increase with age (Grossman). In terms of nutrient intake, this implies that older individuals are more likely to have better diets than younger individuals.

These theoretical models do not indicate where along the intake distribution the predicated effects are likely to occur. The predications are based on the theoretical link between inputs and health. However, the observed input distribution, as in the case of many nutrients, may be such that parts of the distribution are within acceptable risk levels while other parts exceed acceptable risk. In such cases, it is clearly questionable to estimate the predicated effect at one point of the input distribution under the implicit assumption that the effects are same at all points. The more informative approach is to estimate the effects at different points of the distribution so as to find where a predicted effect is the largest. This can be accomplished by estimating a set of quantile regressions for each input of interest.

Although information and health behavior variables act as intermediate pathways through which basic sociodemographic variables influence dietary intakes (Gould and Lin), we excluded them from our estimated functions to avoid potential endogeneity problems. Moreover, our objective was not to estimate the effects of consumers' intermediate choices on dietary intakes, but to map the *net* effect of key sociodemographic variables at different points along the intake distribution.

Data

The nutrient intake data for men and women were obtained from USDA's 1994–96 CSFII (Tippett and Cypel). Each year of the three-year CSFII comprised a nationally representative sample of noninstitutionalized persons residing in the United States. Dietary data for selected sample persons from a screened sample of 9664 households were collected on two nonconsecutive days through in-person interviews using 24-hour recalls. Information on food intakes for both days was provided by 15,303 persons, giving a two-day response rate of 76.1%. From these, we selected adults 18 or above in age and excluded pregnant and lactating women.

By combining the food records with a nutrient database, CSFII provides information on the intakes of a variety of macronutrients, vitamins, and minerals. We used the mean daily intakes of total fat, saturated fat, cholesterol, and fiber as our dependent variables. We focused on these four macronutrients because of their links to cardiovascular disease, obesity, and certain types of cancer (National Research Council). Recommendations on their intake are specifically mentioned in the Dietary Guidelines for Americans.

Table 3 lists the explanatory variables used in the regressions. These sociodemographic and anthropometric variables fall into two sets: household characteristics and personal characteristics. Additionally, we included several survey variables to control for time-related variations in nutrient intakes and a variable to indicate whether the respondent was on any type of special diet. Income, household size, education level, age, height, and weight are continuous variables. The remaining variables are dummy indicator variables. This basic set of sociodemographic variables has been used in most previous studies of nutrient intake. Income is the gross household income in the previous calendar year from all sources before taxes. Height and weight were included to control for the influence of body mass on the amount of food intake. While previous studies have often used the body mass index (BMI) for this purpose, estimating a single coefficient for BMI implies a restriction on the coefficients for BMI components, height and weight. Therefore, we left height and weight in the unrestricted form. Not listed in table 3, but included in all regressions, were fifteen dummy variables representing time effects and special diet status. These were two for survey year, six for survey season, two indicating whether the intake was recorded on a weekend day, four indicating respondent's opinion whether the recorded intake was less than or more than his or her usual intake, and one indicating whether the respondent was on any kind of diet. After dropping observations that were incomplete with respect to the explanatory variables, the final sample sizes were 4725 for men and 4362 for women.

Table 3. Explanatory Variables, Means, and Sample Size

Variable	Men	Women
Household characteristics		
Gross annual income ($',000)	39.5	36.1
Household size	2.9	2.8
Region (Northeast omitted)		
Midwest	0.24	0.25
South	0.36	0.37
West	0.21	0.20
Urbanization (city omitted)		
Suburb	0.46	0.44
Nonmetro	0.26	0.25
Personal characteristics		
Level of education (years)	12.7	12.6
Age (years)	49.0	49.1
Height (inches)	69.8	64.0
Weight (pounds)	183.4	151.7
Race (White omitted):		
Black	0.10	0.13
Asian	0.02	0.02
Other[a]	0.05	0.04
Ethnic origin Hispanic	0.08	0.08
Sample size	4725	4362

Note: Additionally, two dummy variables representing survey year, six representing survey season, two indicating whether the intake was recorded on a weekend day, four indicating the respondent's opinion whether the recorded intake was less than or more than the usual intake, and one indicating whether the respondent was on any kind of diet were used in all regressions.

[a] American Indian, Alaskan native, or other.

Based on the theoretical models discussed earlier, we expect education and age to have negative effects on the intake of fats and cholesterol and positive effects on the intake of fiber. For income, the effects are harder to predict given that inputs such as fats have both positive (taste) and negative (health risk) attributes. Income may increase consumption of fat-rich foods with taste enhancing attributes while income may also provide better access to health information tending to discourage consumption of unhealthful foods. The net effect of income will depend on which of these effects is dominant. If the informational effect is dominant, then income will have an effect similar to the effect of education.

Quantile Regression

Koenker and Bassett introduced quantile regression as a generalization of the sample quantiles to conditional quantiles expressed as linear functions of explanatory variables. This is analogous to the OLS regression model that expresses the conditional mean in a linear form. However, by permitting conditional functions to be specified at *any* quantile, quantile regression enables one to describe the entire conditional distribution of the dependent variable given a set of regressors. The familiar least absolute deviation estimator is a special case of quantile regression that expresses the conditional median as a linear function of covariates. Quantile regression's ability to characterize the whole conditional distribution is most potent when there is heteroskedasticity in the data (Deaton). When the data are homoskedastic, the sets of slope parameters of conditional quantile functions at each point of the dependent variable's distribution will be identical with each other and with the slope parameters of the conditional mean function. In this case, the quantile regression at any point along the distribution of the dependent variable reproduces the OLS slope coefficients, and only the intercepts will differ.

However, when the data are heteroskedastic (that is, the conditional variance of a dependent variable's distribution is not constant but differs by the level of independent variables), the sets of slope parameters of the conditional quantile functions will differ from each other as well as from the OLS slope parameters. Therefore, estimating conditional quantiles at various points of the distribution of the dependent variable will allow us to trace out different marginal responses of the dependent variable to changes in the explanatory variables at these points.

Two additional features of the quantile regression model are noteworthy (Buchinsky 1998). First, the classical properties of efficiency and minimum variance of the least squares estimator are obtained under the restrictive assumption of independently, identically, and normally distributed (i.i.d.) errors. When the distribution of errors is nonnormal, the quantile regression estimator may be more efficient than the least squares estimator. Second, since the objective function for the quantile regression estimator is a weighted sum of absolute deviations, the parameter estimates are robust to outliers.

Estimation and Testing

Let y_i denote intake of a nutrient of the ith sample person, $i = 1, \ldots, N$. We assume that the θth quantile $(0 < \theta < 1)$ of the conditional distribution of y_i is linear in a $K \times 1$ vector of explanatory variables, $\mathbf{x}_i$:

$$(1) \qquad Q_\theta(y_i \mid \mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}_\theta$$

where $Q_\theta(y_i \mid \mathbf{x}_i)$ is the conditional quantile function and $\boldsymbol{\beta}_\theta$ is the unknown vector of parameters. The quantile regression estimator of $\boldsymbol{\beta}_\theta$ is obtained by solving

$$(2) \qquad \min_{\boldsymbol{\beta}_\theta} \frac{1}{N}\left\{ \sum_{i:y_i \geq \mathbf{x}_i'\boldsymbol{\beta}_\theta} \theta|y_i - \mathbf{x}_i'\boldsymbol{\beta}_\theta| + \sum_{i:y_i < \mathbf{x}_i'\boldsymbol{\beta}_\theta} (1-\theta)|y_i - \mathbf{x}_i'\boldsymbol{\beta}_\theta| \right\}.$$

This minimization problem has a linear programming representation, which is guaranteed to have a solution in a finite number of simplex iterations (Buchinsky 1998). Several estimators for the asymptotic covariance matrix for $\hat{\boldsymbol{\beta}}_\theta$ obtained from the above minimization are available, but for obvious reasons, those that rely on the assumption of i.i.d. errors are of limited value (Deaton). Buchinsky (1995) has shown that the design matrix bootstrap estimator provides a consistent estimator for the covariance matrix under very general conditions. In the design matrix bootstrap, quantile regression is estimated on a sample of N observations $(y_i^*, \mathbf{x}_i^*)$, $i = 1, \ldots, N$, drawn randomly, with replacement, from the original sample. The process is repeated B times to obtain bootstrap estimates $\hat{\boldsymbol{\beta}}^*_{\theta b}$, $b = 1, \ldots, B$. The covariance matrix of $\hat{\boldsymbol{\beta}}_\theta$ is then obtained as the covariance of $\hat{\boldsymbol{\beta}}^*_\theta$ computed from the B bootstrap estimates with $\hat{\boldsymbol{\beta}}_\theta$ as the pivotal value.

The minimum-distance method can be used to test for the equality of slope coefficients of a given dependent variable across all estimated quantiles (Buchinsky 1998). Let $\hat{\boldsymbol{\beta}}_P = (\hat{\boldsymbol{\beta}}'_{\theta_1}, \ldots, \hat{\boldsymbol{\beta}}'_{\theta_P})'$ be a $KP \times 1$ stacked vector of unrestricted parameter estimates from quantile regressions at P quantiles. Let $\boldsymbol{\beta}^R = (\beta_{\theta_1 1}, \ldots, \beta_{\theta_P 1}, \beta_2, \ldots, \beta_K)'$ be a $(K+P-1) \times 1$ vector comprising P unrestricted intercepts and $K-1$ restricted slope parameters. The restricted parameter vector $\boldsymbol{\beta}^R$ is obtained by minimizing

$$(3) \qquad Q(\boldsymbol{\beta}^R) = (\hat{\boldsymbol{\beta}}_P - \mathbf{R}\boldsymbol{\beta}^R)'\mathbf{A}^{-1}(\hat{\boldsymbol{\beta}}_P - \mathbf{R}\boldsymbol{\beta}^R)$$

where $\mathbf{A}$ is a positive definite matrix and $\mathbf{R}$ is the appropriate restriction matrix. Under the null hypothesis of the equality of slope coefficients, $NQ(\boldsymbol{\beta}^R)$ is distributed χ^2 with $(PK - P - K + 1)$ degrees of freedom. Since the equality of slope parameters will hold if the i.i.d. assumption is valid, this is a general test for heteroskedasticity. The optimal choice for $\mathbf{A}$ is the variance–covariance matrix of $\hat{\boldsymbol{\beta}}_P$, denoted by $\hat{\boldsymbol{\Lambda}}_P$. Given $\hat{\boldsymbol{\Lambda}}_P$, the usual F-statistic for testing linear restrictions can be used to test for the equality of the slope parameters for a specific explanatory variable at symmetrical quantiles such as $0.1q$ and $0.9q$. If the null hypothesis of homoskedasticity or the equality of the slope coefficients is not rejected, the restricted slope estimates $\boldsymbol{\beta}^R$ give an optimal combination of the quantile slope estimates. Also, given $\hat{\boldsymbol{\Lambda}}_P$, the variance–covariance matrix of the restricted parameter vector can be obtained as $\hat{\boldsymbol{\Lambda}}^R = (\mathbf{R}'\hat{\boldsymbol{\Lambda}}_P^{-1}\mathbf{R})^{-1}$.

The quantile regression parameter estimates are obtained by estimating a separate equation at each quantile of each macronutrient. The variance–covariance matrix of the estimates can be obtained by bootstrapping each of these equations separately. However, to carry out tests of the equality of slope coefficients for a given dependent variable across the P estimated quantiles and to obtain the restricted parameter estimates and their standard errors, it is necessary to have $\hat{\boldsymbol{\Lambda}}_P$, the variance–covariance matrix of the stacked vector of parameter estimates at the P quantiles. This can be obtained by simultaneously estimating quantile regressions at the P quantiles for each bootstrap sample. Thus, the following procedure was used for the estimation and testing of the quantile regressions for each macronutrient. First, the coefficient estimates for a macronutrient were obtained by running quantile regressions separately at the P desired quantiles. Second, a bootstrap sample was drawn for that macronutrient and the bootstrap estimates for the P quantiles were obtained by running quantile regressions separately at the P quantiles for that sample. Finally, after repeating the bootstrap procedure B times, $\hat{\boldsymbol{\Lambda}}_P$ was calculated to obtain the standard errors of the coefficient estimates and to conduct the equality tests. This estimation process was carried out in Stata using the sqreg procedure (Gould 1997). Additional details regarding the estimation of the quantile regression model and the asymptotic covariance matrix of the parameters are in Buchinsky's (1998) methodological survey.

Results

Conditional quantile functions for the intake of the four macronutrients were estimated separately for men and women at five selected quantiles ($P = 5$). In tables 4 to 6 we report coefficient estimates for saturated fat, cholesterol, and fiber. The patterns of results for total fat and saturated fat were almost identical. Therefore, we do not present a table for total fat but discuss the results below. Further, the tables report only the estimates for five key variables of policy interest—income, education, age, race (black compared with white), and ethnic origin (Hispanic compared with non-Hispanic). For comparison with the quantile estimates, the second column in each table presents the OLS estimates. We also computed the restricted coefficient estimates using the minimum-distance estimator. To conserve space, we do not present these estimates in the tables but include the key results in the discussion below. The complete set of results for this study is available from the authors upon request. The standard errors for the quantile regression estimates were obtained using the design matrix bootstrap with $B = 500$ replications. The standard errors for the OLS estimates were computed using White's method.

The R^2's are generally low, indicating the high variation in the mean two-day intakes but are in line with previous studies (e.g., Adelaja, Nayga, and Lauderbach). The F-tests for the OLS regressions showed high significance levels in all cases. Table 7 presents two sets of statistics for testing the equality of slope coefficients across the quantiles. One set gives F-statistics and the associated p-values for the test that, for a given independent variable, its slope at the 0.9 quantile is equal to its slope at the 0.1 quantile ($q_{10} = q_{90}$) and its slope at the 0.75 quantile is equal to its slope at the 0.25 quantile ($q_{25} = q_{75}$). As noted earlier, if an F-test does not reject the equality of slopes at symmetrical quantiles, then the restricted coefficient estimate β^R gives the optimal combination of the five quantile slope coefficients. Such estimates, in general, have lower variance than least squares estimates (Buchinsky 1998). Comparing the

Table 4. Quantile Regression Estimates: Saturated Fat Intake, 1994–96

		Quantile				
Variable	OLS	0.10	0.25	0.50	0.75	0.90
Men						
Income ($\times 10^{-3}$)	−0.016	−0.009	−0.005	−0.011	−0.016	−0.019
	(1.73)	(0.94)	(0.53)	(1.11)	(1.09)	(0.84)
Education	−0.177	0.053	0.004	−0.154	−0.264	−0.521
	(2.36)	(0.65)	(0.05)	(1.74)	(2.13)	(2.56)
Age	−0.243	−0.126	−0.169	−0.222	−0.277	−0.363
	(17.01)	(9.15)	(13.98)	(15.49)	(13.11)	(10.60)
Black	−2.154	−1.182	−1.501	−1.967	−2.687	−2.046
	(2.85)	(1.54)	(2.28)	(2.59)	(2.54)	(1.13)
Hispanic	−2.464	−1.507	−0.747	−2.735	−3.756	−5.635
	(2.88)	(1.72)	(0.98)	(2.57)	(2.35)	(2.51)
R^2	0.153	0.078	0.084	0.087	0.098	0.105
Women						
Income ($\times 10^{-3}$)	−0.010	0.010	0.001	−0.012	−0.009	−0.018
	(1.47)	(1.64)	(0.07)	(1.68)	(0.84)	(1.11)
Education	0.035	0.039	0.042	0.047	0.015	0.035
	(0.64)	(0.74)	(0.65)	(0.70)	(0.17)	(0.26)
Age	−0.101	−0.036	−0.056	−0.096	−0.125	−0.150
	(10.47)	(3.65)	(5.01)	(9.23)	(7.72)	(6.58)
Black	−0.057	−0.333	0.099	0.577	−0.632	0.328
	(0.11)	(0.64)	(0.18)	(1.16)	(0.85)	(0.27)
Hispanic	−1.459	−0.704	−1.118	−1.503	−1.291	−1.561
	(2.29)	(0.99)	(2.00)	(1.98)	(1.25)	(1.21)
R^2	0.104	0.049	0.050	0.061	0.065	0.079

Note: Absolute t-values reported in parentheses. All regressions included an intercept and 25 additional explanatory variables; see table 3 for definitions.

Table 5. Quantile Regression Estimates: Cholesterol Intake, 1994–96

		Quantile				
Variable	OLS	0.10	0.25	0.50	0.75	0.90
Men						
Income ($\times 10^{-3}$)	−0.388	−0.060	0.052	−0.315	−0.615	−0.835
	(3.13)	(0.62)	(0.47)	(2.60)	(3.46)	(2.70)
Education	−5.440	0.411	−1.830	−5.195	−9.945	−7.131
	(5.19)	(0.48)	(2.05)	(4.65)	(5.49)	(2.80)
Age	−1.306	−0.415	−0.807	−1.174	−1.215	−2.357
	(7.09)	(2.95)	(4.68)	(5.59)	(4.36)	(5.21)
Black	56.514	16.259	36.393	46.731	89.803	115.965
	(4.86)	(1.85)	(3.43)	(3.36)	(3.50)	(3.72)
Hispanic	−6.454	−1.759	−2.825	−0.361	10.891	25.56
	(0.47)	(0.18)	(0.25)	(0.02)	(0.48)	(0.77)
R^2	0.084	0.035	0.035	0.047	0.062	0.065
Women						
Income ($\times 10^{-3}$)	−0.217	0.022	0.067	−0.108	−0.445	−0.363
	(2.44)	(0.31)	(0.84)	(1.24)	(2.75)	(1.55)
Education	−1.458	−0.261	−1.215	−1.366	−1.271	−3.508
	(1.84)	(0.33)	(1.80)	(1.72)	(0.97)	(1.58)
Age	−0.179	−0.032	−0.138	−0.114	−0.045	−0.073
	(1.29)	(0.29)	(1.28)	(0.80)	(0.19)	(0.18)
Black	36.176	13.846	25.535	35.149	58.40	63.415
	(4.88)	(2.70)	(4.14)	(4.16)	(4.96)	(3.18)
Hispanic	−3.366	5.829	6.473	2.241	−2.314	−14.023
	(0.39)	(0.79)	(0.86)	(0.20)	(0.18)	(0.61)
R^2	0.060	0.030	0.030	0.034	0.046	0.051

Note: Absolute t-values reported in parentheses. All regressions included an intercept and 25 additional explanatory variables; see table 3 for definitions.

Table 6. Quantile Regression Estimates: Fiber Intake, 1994–96

Variable	OLS	Quantile 0.10	0.25	0.50	0.75	0.90
Men						
Income ($\times 10^{-3}$)	0.024	0.017	0.020	0.021	0.035	0.040
	(4.08)	(2.82)	(3.19)	(3.51)	(4.12)	(2.75)
Education	0.239	0.196	0.291	0.292	0.284	0.303
	(4.47)	(3.77)	(5.34)	(5.61)	(3.81)	(2.22)
Age	−0.004	0.007	0.010	0.007	−0.013	−0.037
	(0.50)	(0.96)	(1.16)	(0.78)	(1.00)	(1.96)
Black	−2.188	−1.465	−1.444	−2.300	−2.780	−2.073
	(4.97)	(3.16)	(3.58)	(5.65)	(3.47)	(1.80)
Hispanic	0.789	0.586	0.173	0.767	1.511	0.984
	(1.39)	(1.01)	(.30)	(1.14)	(1.54)	(0.64)
R^2	0.078	0.049	0.045	0.050	0.054	0.057
Women						
Income ($\times 10^{-3}$)	0.023	0.016	0.020	0.026	0.020	0.039
	(5.18)	(4.45)	(5.00)	(4.84)	(3.05)	(3.65)
Education	0.351	0.253	0.270	0.387	0.470	0.354
	(8.54)	(6.15)	(6.76)	(8.15)	(6.94)	(4.79)
Age	0.035	0.029	0.030	0.042	0.048	0.035
	(5.18)	(4.48)	(4.80)	(5.09)	(4.80)	(1.96)
Black	−0.847	−0.682	−1.195	−1.069	−1.206	−0.660
	(2.61)	(2.37)	(3.84)	(2.92)	(2.18)	(0.82)
Hispanic	0.836	0.995	0.918	1.047	1.036	0.763
	(1.98)	(2.29)	(2.48)	(2.23)	(1.49)	(0.76)
R^2	0.104	0.063	0.067	0.066	0.069	0.067

Note: Absolute t-values reported in parentheses. All regressions included an intercept and 25 additional explanatory variables; see table 3 for definitions.

restricted estimates with the corresponding OLS estimates, we found that in almost all cases β^R was more precisely estimated with lower standard errors than the corresponding OLS estimates.

The second set of statistics in table 7, reported under each dependent variable, gives the χ^2 values for the test that all slope parameters for that dependent variable are equal across the five quantiles. Since $K = 31$ and $P = 5$, these χ^2 statistics have 120 degrees of freedom. In all cases, the equalities of slope parameters across the five quantiles are rejected at $p < 0.001$. The χ^2 tests, therefore, suggest that there is significant heteroskedasticity in the nutrient intake data.

For both men and women, household income did not significantly influence saturated fat intake (and total fat intake) at any quantile. The income coefficients for the fat intakes were also insignificant under OLS. However, for cholesterol, the additional information revealed by the quantile estimates as compared with the OLS estimates comes into sharper focus (table 5). The OLS estimates showed that income had a negative (and healthwise, beneficial) effect on cholesterol intake of both men and women. The quantile estimates showed that much of this beneficial effect was located at the upper quantiles. At these quantiles in the observed distribution, the cholesterol intakes exceed the recommended level (table 1). For men, the effect of income on cholesterol intake at $0.9q$ was 115% larger than the OLS estimate. This implies that, holding other variables constant, as income increases the upper conditional quantiles of cholesterol intake are decreasing more rapidly than is the conditional mean. For men and women, the equality restrictions on income coefficients across symmetrical quantiles ($q_{10} = q_{90}$ and $q_{25} = q_{75}$) were rejected at the 10% level.

Income had a positive effect on fiber intake at all quantiles. However, for men and women, the largest effect was at $0.9q$. Since the dietary risk for fiber (inadequacy) is relatively greater at the lower end of the intake distribution, the largest effect for income might be expected at the bottom quantiles. This was not the case, perhaps because except for a few age groups, most adults had inade-

Table 7. Tests for Equality of Slope Parameters across Quantiles

Variable	Saturated Fat $q_{10} = q_{90}$	Saturated Fat $q_{25} = q_{75}$	Cholesterol $q_{10} = q_{90}$	Cholesterol $q_{25} = q_{75}$	Fiber $q_{10} = q_{90}$	Fiber $q_{25} = q_{75}$
Men						
Income	0.19	0.57	6.02	12.45	2.34	2.99
	(0.66)	(0.45)	(0.01)	(0.00)	(0.13)	(0.08)
Education	7.44	4.86	8.76	20.41	0.59	0.01
	(0.01)	(0.03)	(0.00)	(0.00)	(0.44)	(0.93)
Age	45.26	27.14	20.15	2.16	4.63	2.82
	(0.00)	(0.00)	(0.00)	(0.14)	(0.03)	(0.09)
Black	0.21	1.29	10.25	4.81	0.27	2.99
	(0.65)	(0.26)	(0.00)	(0.03)	(0.60)	(0.08)
Hispanic	3.08	3.97	0.67	0.41	0.07	2.01
	(0.08)	(0.05)	(0.41)	(0.52)	(0.80)	(0.16)
$\chi^2(120)$	370.28		430.79		271.45	
Women						
Income	2.83	0.74	2.62	10.26	4.49	0.00
	(0.09)	(0.39)	(0.11)	(0.00)	(0.03)	(0.94)
Education	0.00	0.09	2.15	0.00	1.66	10.27
	(0.98)	(0.76)	(0.14)	(0.97)	(0.20)	(0.00)
Age	23.44	17.68	0.01	0.17	0.12	3.43
	(0.00)	(0.00)	(0.92)	(0.68)	(0.73)	(0.06)
Black	0.27	0.88	6.06	8.07	0.00	0.00
	(0.60)	(0.35)	(0.01)	(0.00)	(0.98)	(0.98)
Hispanic	0.36	0.03	0.72	0.46	0.05	0.03
	(0.55)	(0.86)	(0.40)	(0.50)	(0.82)	(0.86)
$\chi^2(120)$	261.37		281.00		216.06	

Note: The numbers against the variable names are F-statistics with $(1, N-K)$ degrees of freedom. The associated p-values are reported in parentheses. The χ^2 statistics for the restriction that all slope parameters for a given dependent variable are equal across the five quantiles.

quate fiber intake even at the 90th percentile (see tables 1 and 2).

The effect of educational attainment on intakes of men clearly illustrates the importance of examining the whole conditional distribution rather than just the conditional mean. An additional year of education reduced men's saturated fat intake by 0.18 gram at the conditional mean (table 4). However, at $0.9q$, an additional year of education reduced saturated fat intake by 0.52 gram, nearly a 200% increase in the estimated effect. Both for saturated fat and total fat, quantile estimates at the median and below were insignificant. In fact, for total fat, the marginal effect of education at $0.1q$ was positive and numerically large (0.32 gram). This is not surprising since at $0.1q$, the observed intakes were substantially below the recommended upper levels.

Men's intake of cholesterol was influenced by their educational attainment in a similar fashion. The reduction in intakes attributable to education was larger at the upper conditional quantiles compared with the conditional mean. For fiber, education had a more uniform effect across the quantiles and the equality of coefficients at symmetrical quantiles could not be rejected. As with income, the lack of a larger effect at the bottom quantiles may be because, for the most part, the entire observed distribution of fiber intake is below the recommended fiber intake.

The results for men confirm that education is positively correlated with better diets, just as it has been shown to be positively correlated with other desirable health behaviors (Grossman and Kaestner). However, for fats and cholesterol, our results show something new. The beneficial effects of education are much greater where they matter most—at the upper quantiles where the risk of excess intakes is greater. For fiber, although the education coefficients tended to be larger at the upper quantiles and not at the bottom quantiles, the size of the coefficients was relatively similar across the quantiles.

There is less evidence of an increasing marginal effect of education at the higher quantiles for women. Except for fiber, the OLS and quantile estimates for education were insignificant. For cholesterol, although the quantile estimates were insignificant, their optimal combination—the restricted estimate—was negative and significant (−2.4, t-value = −5.5). The expected effect of education may not materialize as strongly for women because their risk of excess intake is less than men's. For example, women's cholesterol intake exceeds the recommended level only at the 90th percentile, whereas men's intake is above the limit at the 75th percentile. For fiber, similar to men, the effect of women's educational attainment was significant at all quantiles.[1]

Consistent with the prediction from Grossman's model, men's total fat, saturated fat, and cholesterol intakes declined with age, but more notably, the rate of decline rose steadily from the lower to the upper quantiles. The differences in the estimated effects between $0.1q$ and $0.9q$ were over 200%, and between the OLS and $0.9q$ over 50%, for these macronutrients. The strong intake response to age is not surprising given that the health risk of poor diets is cumulative and increases with age. Therefore, older individuals will display a greater propensity to improve their diets compared with younger individuals. In addition, our results show that the age effect is strongest at the riskier part of the intake distribution.

Women's age had a similar pattern of impact on saturated fat intakes. For cholesterol, the women's age coefficient was insignificant under OLS and at all quantiles. However, the optimally combined restricted estimate was significant (−0.18, t-value = −2.49), showing that cholesterol intake among women does decline with age. The age effect on women's fiber intake was uniformly positive across the OLS and quantile estimates. The age effect on fiber intake of men showed conflicting results under the different estimators. While the OLS and lower quantile estimates were insignificant, the $0.9q$ estimate was negative and significant (table 6). Although the restricted estimate was positive and significant in accordance with our expectation, the numerical effect was small (0.013, t-value = 2.52). It is not clear why age, which had sizable effects on other macronutrient intakes, would have such limited impact on men's fiber intake.

The difference in saturated fat intake between black and white men tended to increase with the quantiles, although interquantile equality could not be rejected. The restricted estimate showed that black men had significantly lower saturated fat intake compared with white men (−1.43, t-value = −3.06). While black men's intakes looked better in terms of saturated fat (and total fat as well; restricted estimate = −3.33, t-value = −2.42), the picture was starkly different for cholesterol and fiber. After controlling for other effects, black men had higher cholesterol intake and lower fiber intake relative to white men. The point estimate at $0.9q$ showed black men consuming 116 milligrams more cholesterol than white men. This is twice what the OLS estimate would have led one to believe. The black–white difference in fiber intake was 2.3 grams at the median (that is, 0.5q). Given that the median fiber intake is 16.2 grams, this is a sizable difference (14%).

Similar to the black–white difference for men, black women had higher cholesterol intake and lower fiber intake compared with white women. The black–white difference in cholesterol was large, particularly at the upper quantiles, and the interquantile equality tests showed significant difference between intakes at the lower and upper quantiles. The black–white difference in fiber intake for women was 1 gram at the conditional median. The restricted estimate, although of similar size, was much more precise (−1.05, t-value = −5.50). Unlike black men though, black women's intakes of total and saturated fats were not significantly lower than white women's.

Hispanic men's diets were significantly lower in saturated fat intake compared with the diets of non-Hispanic men. Based on the restricted quantile estimate, Hispanic men consumed 1.7 grams less saturated fat compared with non-Hispanic men. The quantile estimates for saturated fat indicate that the relative difference was located at the upper ends of the distribution. This was confirmed by interquantile equality tests, which were both rejected at the 10% level. Due to the relatively low t-values, no significant difference between Hispanic and non-Hispanic groups was evident for cholesterol and fiber intakes at various quantiles. However, the restricted estimate did show a higher fiber intake for Hispanic men (0.85, t-value = 2.25).

---

[1] In this study, we treated education as a continuous variable. While this offers the convenience of interpreting a single coefficient, it is possible that the effect of education may vary by the level of educational attainment. This can be investigated by entering the education variable in a nonlinear form or by using a spline function. We leave this for future research.

Among women, saturated fat intake was about 0.8 gram lower for Hispanics compared with non-Hispanics, based on the restricted estimate (−0.82, t-value = −2.13). The quantile estimates for women's cholesterol intake showed an interesting trend, with the Hispanic/non-Hispanic difference reversing in sign from lower to upper quantiles, although none of the estimates were significant. Hispanic women's fiber intake was significantly higher at the bottom end of the distribution. At $0.1q$, the unconditional estimate of which is well below the recommended level, Hispanic women tended to consume about 1 gram of fiber more than non-Hispanic women. Overall, the results for men and women show that the Hispanic segment of the U.S. adult population tends to have a healthier macronutrient intake profile than non-Hispanics.

The data tables compiled from the 1994–96 CSFII can be used to speculate on the likely food sources of the nutrient intake differences by race and ethnicity (U.S. Department of Agriculture 1998, 1999). For example, black men over age 20 consume only 7 grams of cheese, compared with 20 grams consumed by white men over age 20. The most recent data on the sources of nutrients (Subar et al.) show that cheese is the largest source of saturated fat and the fourth largest source of total fat among U.S. adults.[2] Black men also tend to consume less whole and low-fat milk, another major source of fats, compared with white men. However, black men consume less yeast breads and ready-to-eat cereals, and more eggs compared with white men, which may account for their lower fiber intake and higher cholesterol intake.

The picture is less clear regarding the sources of the Hispanic/non-Hispanic difference. For example, Hispanic men over age 20 consume more beef and about the same amount of cheese as non-Hispanic white men over age 20. Consequently, the difference in their total and saturated fat intakes must come from other sources. However, these reported food intakes are mean intakes. A clearer answer about the sources of intake differences may require a comparison of the distribution of food intakes at different percentiles, similar to those reported in table 1 for the macronutrients.

Conclusion

Understanding and quantifying the relative differences in food and nutrient intakes among population subgroups is important for guiding nutrition promotion expenditures. The results can also contribute to improved understanding of health risk behavior at a time of rapidly evolving health information environment. However, the nature of intake distributions is such that the risk of dietary excess or inadequacy is greater at the tails than at the mean. Consequently, it is questionable to assume that the marginal effects of population characteristics will be constant along all parts of the conditional distribution of intakes. In this case, any analysis that focuses on only one part of the distribution, such as the conditional mean, may give an incomplete picture of the sources of intake differences in the population. In this study, we used quantile regression, a method suited for characterizing the entire distribution of intake, to examine macronutrient intakes among U.S. adults.

The findings clearly suggest that the marginal effects at the tails of the intake distribution are often quite different from those at the mean. A more complete picture of intake differences among population subgroups emerges from the quantile regression estimates than from the OLS estimates alone. Of course, this entails a larger number of estimates to consider. But in most cases, an optimal combination of the quantile estimates outperformed the OLS estimates in precision.

These results have important implications for future studies evaluating the dietary impact of nutrition-related policy interventions such as food assistance programs and food labeling regulations. For such studies to fully uncover the extent and nature of the behavioral impact, they must look beyond the conditional mean to parts of the dietary intake distribution where the risk of inadequacy or excess is greatest.

Our results suggest that individuals, particularly men, at higher education and income levels may have benefited more from health and nutrition information initiatives such as

[2] The food intakes reported in the data tables are mean amounts based on day-1 of the 1994–96 CSFII (U.S. Department of Agriculture 1998, 1999). The dietary sources of nutrients are based on day-1 of the 1989–91 CSFII and are not separated by race or sex (Subar et al.). These comparisons are meant to be illustrative, especially considering that the estimated differences reported in this article are conditional *(ceteris paribus)*, whereas the figures from the data tables and the dietary sources study are unconditional.

the Nutrition Labeling and Education Act. Certainly, this explanation is consistent with the effect of human capital on health behavior predicted by the household production model. By comparing the influence of these variables over a sufficiently long time-span, it may be possible to verify the validity of this potentially important linkage.

[Received June 2000; accepted September 2001.]

References

Adelaja, A., R.M. Nayga Jr., and T. Lauderbach. "Income and Racial Differentials in Selected Nutrient Intakes." *Amer. J. Agr. Econ.* 79(December 1997):1452–60.

Bowman, S.A., M. Lino, S.A. Gerrior, and P.P. Basiotis. *The Healthy Eating Index: 1994–96*. Washington DC: U.S. Department of Agriculture, Center for Nutrition Policy and Promotion, CNPP-5, July 1998.

Buchinsky, M. "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models: A Monte Carlo Study." *J. Econometrics* 68(August 1995):303–38.

——. "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research." *J. Human Resour*. 33(Winter 1998):88–126.

Chavas, J.-P., and K.O. Keplinger. "Impact of Domestic Food Programs on Nutrient Intake of Low-Income Persons in the United States." *S. J. Agr. Econ*. 15(July 1983):155–63.

Chern, W.S. "Demand for Food and Demand for Health in the United States." Paper presented at the mini-symposium on "Effects of Health Information on the Demand for Food—EU and US Experiences" at the XXIV International Conference of the International Association of Agricultural Economists, Berlin, Germany, 13–18, August 2000.

Davis, C., and E. Saltos. "Dietary Recommendations and How They Have Changed Over Time." *America's Eating Habits: Changes and Consequences*. E. Frazao, ed., pp. 33–50. Washington DC: U.S. Department of Agriculture, Economic Research Service, Agricultural Information Bulletin No. 750, 1999.

Deaton, A. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Baltimore MD: Johns Hopkins University Press, 1997.

Gould, B.W. "Factors Affecting U.S. Demand for Reduced-Fat Fluid Milk." *J. Agr. Resour. Econ.* 21(1996):68–81.

Gould, B.W., and H.C. Lin. "Nutrition information and household dietary fat intake." *J. Agr. Resour. Econ*. 19(December 1994):349–65.

Gould, W. "sg70: Interquantile and Simultaneous-Quantile Regression." *Stata Tech. Bull.* 38(July 1997):14–22.

Grossman, M. "On the Concept of Health Capital and the Demand for Health." *J. Polit. Econ*. 80(March/April 1972):223–55.

Grossman, M., and R. Kaestner. "Effects of Education on Health." *The Social Benefits of Education*. J.R. Behrman and N.G. Stacey, eds., pp. 69–123. Ann Arbor MI: University of Michigan Press, 1997.

Ippolito, P.M., and A.D. Mathios. "Information, Advertising, and Health Choices: A Study of the Cereal Market." *Rand J. Econ*. 21(Autumn 1990):459–80.

Kantor, L.S. *A Dietary Assessment of the U.S. Food Supply: Comparing per Capita Food Consumption with Food Guide Pyramid Serving Recommendations*. U.S. Department of Agriculture, Economic Research Service, Agricultural Economic Report No. 772, 1998.

Koenker, R., and G. Bassett. "Regression Quantiles." *Econometrica* 46(January 1978):33–50.

Lancaster, K. *Consumer Demand: A New Approach*. New York: Columbia University Press, 1971.

Lin, B.-H., J. Guthrie, and E. Frazao. *Away-From-Home Foods Increasingly Important to Quality of American Diet*. U.S. Department of Agriculture, Economic Research Service, Agricultural Information Bulletin No. 749, 1999.

National Research Council. *Diet and Health: Implications for Reducing Chronic Disease Risk*. Washington DC: National Academy Press, 1989.

Nayga, R.M., Jr. "Effects of Socioeconomic and Demographic Factors on Consumption of Selected Food Nutrients." *Agr. Resour. Econ. Rev.* 23(October 1994):171–82.

Strauss, J., and D. Thomas. "Human Resources: Empirical Modeling of Household and Family Decision." *Handbook of Development Economics, Vol. III.* T.N. Srinivasan and J.R. Behrman, eds. Amsterdam: North-Holland, 1996.

Subar, A.F., S.M. Krebs-Smith, A. Cook, and L. Kahle. "Dietary Sources of Nutrients Among US Adults, 1989 to 1991." *J. Amer. Diet. Assoc.* 98(May 1998):537–547.

Tippett, K.S., and Y.S. Cypel, eds. *Design and Operation: The Continuing Survey of Food*

Intakes by Individuals and the Diet and Health Knowledge Survey, 1994–96. Beltsville MD: U.S. Department of Agriculture, Agricultural Research Service, NFS Report No. 96-1, 1998.

U.S. Department of Agriculture, Agricultural Research Service. *Data Tables: Food and Nutrient Intakes by Race, 1994–96.* Online on the "Products" page at http://www.barc.usda.gov/bhnrc/foodsurvey/home.htm, August 1998.

——. *Data Tables: Food and Nutrient Intakes by Hispanic Origin and Race, 1994–96*. Online on the "Products" page at http://www.barc.usda.gov/bhnrc/foodsurvey/home.htm, July 1999.